

驾驶行为对车辆风险评级的实证分析

单 鹏¹ 华晓东² 邓颖璐²

(1. 中国人民大学财政金融学院, 北京 100872; 2. 清华大学经济管理学院, 北京 100084)

[摘 要] 本文主要探究了一些反映驾驶行为的车联网因子在车辆风险评级中所起到的作用, 分析了在我国推行国际流行的“基于使用的保险”的必要性与可行性。回归分析结果显示, 新加入的动态因子, 包括急加速、速度、油耗等, 确实具有相当显著的解释力, 能够弥补性别、年龄等传统静态因子的不足, 而且还可能实现保费的减少。在此基础上, 尝试提出了“基础费率+浮动费率”模式, 作为未来的保费定价建议。

[关键词] 车联网; 基于使用的保险; 保费定价

[中图分类号] F840.63 **[文献标识码]** A **[文章编号]** 1004-3306(2017)01-0068-12

DOI: 10.13497/j.cnki.is.2017.01.007

一、引 言

汽车保险是我国财险的重要板块。数据显示, 2014 年我国保险公司的车险保费总收入高达 5515.9 亿元, 在整个财险市场中占比超 7 成, 车险对总承保利润的贡献也接近 80%。^[1] 因此, 车险的创新与发展是我国各保险公司的重要课题。不同的车险发展历阶段, 对应不同的定价模式。早期是逻辑较简单、单一的“保额定价”, 即根据新车购置价确定车损险保额, 再依据保额决定保费。后来有了“车型定价”的概念, 即建立起一套健全的车辆标准体系, 通过归类分析得出不同车型的保险损失情况, 进而完成保费定价。渐渐人们又认识到, 仅仅依靠为数不多的、静态的车辆性质因素进行定价不能足够精确、客观地反映风险状况, 需要增加更丰富的动态因素, “使用定价”也应运而生。这种“基于使用的保险(以下简称 UBI, Usage-Based Insurance)”是通过前端感应系统记录每个个体车辆使用数据, 建模分析并量化其驾驶行为, 进而进行风险评级, 制定保费。这种 UBI 定价模式, 是当前保险行业的前沿。^[2] 我国当前仍处于保额定价与车型定价之间的阶段, 而在国外(美、英、意)发达保险市场, 车险已进入了 UBI 定价的阶段。无疑, 我国需要尽快向国际领先的定价模式靠拢。

新时代保险业存在两大趋势: 第一种是从“以车为本”到“以人为本”。过去仅与车价相关的、缺乏差异化的保费定价是低效的, 会引发优质车辆投保人补贴劣质车辆投保人的不良现象, 甚至会提高车主的道德风险, 造成理赔失控, 这不利于整个行业。而从人出发的、基于使用量的保险, 趋向于按照实际风险进行个性化车险定价, 提高了合理性和公平性。^[3] 第二种趋势则是从“传统”走向“车联网①”。我们知道, 保险产品定价的根据是基于精算技术对承保风险的预测。传统的风险预测, 是在历史理赔数据和抽样调查的基础上, 利用统计分析(加以一定的主观假设)实现的。而今, 由于车联网的出现, 我们能够采集到更庞大、多样、精确的

[作者简介] 单 鹏, 东北财经大学金融学博士, 中国人民大学财政金融学院博士后, 研究方向: 金融学、保险学、大数据挖掘; 华晓东, 清华大学经济管理学院金融系硕士研究生, 研究方向: 保险与风险管理; 邓颖璐, 德克萨斯大学奥斯汀分校商学院金融学博士, 清华大学经济管理学院金融系助理教授, 清华大学车联网大数据与金融保险研究中心主任, 研究方向: 保险与风险管理。

数据,这些数据也将车险业也带入大数据时代。大数据意味着我们能够低成本地获得全量数据(而不只是样本数据),能够把外部数据(而不只是行业内部数据)也引入车辆风险的描绘中,能够利用实时数据(而不只是历史数据)进行更准确的预测。更重要地,我们能够获取维度更丰富的数据,包括里程、速度、加速度、急转弯、油耗、夜间驾驶比例等等——有了新的车联网因子,我们可以从更多的视角刻画风险,进而实现更科学、更接近现实的风险预测。结合两种趋势,我们认为,人们的微观驾驶行为应当成为未来保险公司进行风险评级的核心要素。

本文目的即通过实证研究证明,将驾驶行为纳入风险评级中是一件有意义的事情。期望通过丰富定价因子,可以增加解释力与预测力,推动更精确、合理的车险定价。这种结果意义深远:对驾驶员而言,享受与自己风险水平匹配的保费定价,会促使自己安全驾驶;对保险行业而言,一来能减少信息不对称的问题,减轻索赔损失,二来也能通过提供差异化的费率,吸引优质客户;在长远中对汽车行业而言,还会不断推动汽车安全技术的进步。

二、文献综述

很多研究者认为基于车联网技术的新型保险产品可以优化保险公司的风险组合。Lindberg(2005)提出,按里程付费的费率厘定模式能形成客户自主选择机制,因而保险公司可以吸引低风险客户,进而优化风险组合。Oberholzer(2003)分析了车联网保险在瑞士发展的潜力,认为低风险和低频驾驶者是保险公司理想的目标客户,建议大型保险公司与汽车商合作,创新保单,开拓新市场。^[7] Persson Johan, Granhage David(2004)认为运输公司具有应用车联网技术的内在动力,应用车联网技术搭建的交通通信信息系统能为保险公司提供更多信息,这些信息可以用于动态定价、风险评估以及设计个性化保险产品,最终降低保费,实现合作共赢。^[8]

Lindberg(2005)通过实地实验,利用车载电子数据收集器和GPS技术获得实时驾驶数据,分析出行和驾驶习惯,据此量化个体风险,并设定不同的保费折扣,该研究认为在这种费率模式下驾驶员将减慢驾驶速度。^[9] Coroama(2006)引入一个依据驾驶行为确定费率的系统模型,所用传感设备记录的行驶数据包含汽车里程,还包括了车轮转速、转向角度和横向加速度等,该模型另一个引人之处在于它能够将驾驶员的驾驶实时表现反馈在图形界面上,并计算出这种表现所对应的保费额,借此可以时刻提醒驾驶员培养良好的驾驶习惯,提高驾驶技术。^[10] Kimberley and Rob(2012)和 Ippisch(2010)认为尽管需要大量前期投入,收益前景同样广阔。

驾驶行为理论是我们选择驾驶行为作为定价因子的理论基础。20世纪60至70年代,学术界提出的模型包括四种:描述性模型、信息处理模型、动机模型和情景中介模型。^[4]

随着车联网技术的不断发展,欧美国家前沿的UBI模型也在不断更新。一代UBI为纯里程因子定价的保险,或以燃油量付费的保险。二代UBI车联网保险产品基于驾驶里程、驾驶方式、驾驶路况进行定价。第三代车联网保险产品使用类似驾驶监督或教练方式的技术来改善驾驶员安全水平,以降低保费。^[5] UBI在国外有着相当成功的实践经历。Metromile公司按里程计费,对一年开不到8000英里(约12000公里)的车主而言,一年可以省下大约424美元(约2600元)。Progressive公司请用户进行了Snapshot测试,根据评分得到优惠,最高可达30%。Lloyd Latchford Schemes则是基于车联网的车队安全管理服务,保险公司可以为车队司机建立风险档案,并为他们量身定制训练计划,以完善不同驾驶员的驾驶行为——降低事故率,加固客户关系。Norwich Union则进行基于驾驶行为的车险定价试点,试点表明很多车主喜欢这种形式的车险产品。^[6]

① 车联网(Internet of Vehicles),即汽车物联网,是一种汽车信息服务。它采用道路和车辆上的电子传感装置收集车辆、道路及环境信息,使车与路、车与车、车与人等之间实时联网,实现信息的互联互通和即时共享,从而实现对车、人、路、环境等的智能监控、调度与管理。

各方面的研究似乎都表明,加入驾驶行为因子的保险定价是一件合理且必要的事情,同时具备着可行性。

三、研究方法及模型

众所周知,保费应反映可能发生的变故对投保人所造成的期望损失。我们以保险公司车险保单的历史数据中记录的实际损失额作为保费预测的基础,即将车辆在给定期限内的赔款总金额(Claim Amount)作为模型的被解释变量。

为帮助保险公司对车辆风险进行更精确的界定与评价,需要选取更丰富、有力的解释变量。一方面,我们仍需传统的静态因子,它们也是过去保险公司进行定价的基础。例如,太平洋保险公司就曾经提出“年龄-驾龄-性别”三维系数作为保费定价的核心因子。事实上,静态因子包括了“人”和“车”两个维度:从人角度,除了性别、年龄等,无赔优(NCD)^①系数、是否续保、车主性质等,均影响车辆风险。特别是无赔款优待系数,常作为保险费率中的重要乘数,比如中国平安费率系数表中,其系数值从0.7到1.3不等,跨度很大。从车角度,车型、车辆种类及其使用性质、新车购置价等,也均是车辆风险的潜在成因,不容忽视。另一方面,由于车联网技术迅猛发展,车联网提供的动态因子也势必成为我们的选择对象。如前文所述,行驶里程、速度、油耗、急加(减)速次数、急转弯次数、夜间出行比例等各种数据都可实时获取,而且均和行车的状况紧密相关,具备相当的解释价值,因此车联网动态因子也应成为不可或缺的预测因子。

我们从上文提及的自变量中,基于数据的可获得性和逻辑的合理性,筛选出一些更有意义的变量,代入模型进行后续分析。最终模型为

$$\text{ClaimAmount} = \beta_0 + (\beta_1 \text{ClaimAdjustValue} + \beta_2 \text{MotorType} + \beta_3 \text{Gender} + \beta_4 \text{Age}) + (\beta_5 \text{Mileage} + \beta_6 \text{DailyHours} + \beta_7 \text{DailyAcceleration} + \beta_8 \text{DailyDeceleration} + \beta_9 \text{NightDrive} + \beta_{10} \text{AverageSpeed} + \beta_{11} \text{FuelConsumption}) \quad (1)$$

该回归模型共有11个解释变量,其中包含4个静态因子,7个动态因子。全体变量定义见表1。其中:车辆种类(代码)是哑变量,若车辆为6座以下的小客车,则变量赋值为0;其余车辆(6座以上的大客车,及货车等),变量赋值为1。自然人性别(代码)也是哑变量,若为女性,则变量赋值为0;若为男性,变量赋值为1。夜间出行比例(20-6)是指,根据车联网公司的定义和算法,将晚20点至早6点定为夜间出行时段,这和市交通委的定义方法也比较一致。

变量定义

表1

变量	类型	说明
ClaimAmount	货币	赔偿总金额
ClaimAdjustValue	数值	无赔优调整系数
MotorType	枚举	车辆种类(代码)
Gender	枚举	自然人性别(代码)
Age	数值	年龄
AnnualMileage	数值	年均行驶里程
DailyHours	数值	日均驾驶时长
DailyAcceleration	数值	日均急加速次数
DailyDeceleration	数值	日均急减速次数
NightDrive	数值	夜间出行比例(20-6)
AverageSpeed	数值	平均速度
FuelConsumption	数值	平均百公里油耗

^① 无赔优即我国的无赔款优待(Non-Claim Discount,简称NCD),反映了被保险人过往保险年度的索赔记录。若在上一保险期限内未发生赔付,则无赔优系数较小,续保时可享受保险公司的保费折扣。

四、数据信息

(一)数据来源

赔偿总金额及各个静态因子的数据来源于中国保险信息技术管理有限责任公司(以下简称中国保信)的车险保单大数据库。各个动态因子的数据来源于北京一家车联网科技公司的车联网数据库。

(二)数据处理

通过采用车辆 VIN 码^①进行匹配,我们提取了车联网公司 10300 名用户的对应保单信息,包括承保记录 293642 条,理赔记录 160354 条。由于险种数据信息存在一些问题,我们将每辆车的各个险种进行去重合并,同时考虑到一年的保险期限,对所有记录做了分年处理,2010 年得到 1130 条记录,2011 年有 1627 条,2012 年有 2372 条,2013 年有 3416 条,最近的 2014 年则有 5320 条数据。至此,初步实现了想要的的形式:以车辆为单位,一辆车对应一条承保及索赔记录。

同时对所需自变量字段进行筛选,通过提取加工生成了车联网动态因子字段,包括年均行驶里程、日均驾驶时长、日均急加速次数、日均急减速次数、夜间出行比例、平均速度、平均百公里油耗,共 7 个;对保单数据的无关字段和不良记录进行了剔除,保留的保单字段包括赔偿总金额、无赔优调整系数、车辆种类代码、投保人性别、年龄,共 5 个。

鉴于车联网数据是最新发布的,而且作为最近的一年,2014 年的数据量又明显多于其他年份,因此我们将以 2014 年的记录作为核心数据,代入模型进行回归分析;同时,也将用类似的方法分析 2010~2013 年的数据,互为稳健性检验。

(三)数据基本特征

1. 描述性统计

2014 年数据的描述性统计见表 2。

描述性统计(2014)

表 2

	ClaimAmount	ClaimAdjustValue	MotorType	Gender
平均	2041.959928	0.889979577	0.06554029	0.848124768
中位数	0	0.9	0	1
众数	0	1	0	1
标准差	8800.224284	0.146656008	0.247499764	0.358933233
方差	77443947.45	0.021507985	0.061256133	0.128833066
最小值	0	-0.3	0	0
最大值	359800	2	1	1
	Age	AnnualMileage	DailyHours	DailyAcceleration
平均	37.6043446	16017.08429	2.18812106	1.51627825
中位数	39	13109.5	1.905	0.461538
众数	39	15083	1.64	0
标准差	7.74040105	11899.40234	1.25487909	3.40135855
方差	59.9138084	141595776.1	1.57472152	11.5692400
最小值	19	238	0.08	0
最大值	79	149385	13.61	64.422222

^① VIN 码即 Vehicle Identification Number(车架号),是制造厂为了识别而给一辆车所指定的一组字母和数字组成,具有对车辆的唯一识别性,相当于车辆的“身份证号”。

(续表)

	DailyDeceleration	NightDrive	AverageSpeed	FuelConsumption
平均	1.432585884	0.117631768	26.32598806	10.5354216
中位数	0.842105	0.1	25.41025	10.20865
众数	1	0	21.8545	0
标准差	2.125642292	0.086747251	6.697333452	2.73889034
方差	4.518355155	0.007525085	44.85427536	7.50152030
最小值	0	0	2.585	0
最大值	43	1	83.3747	39.6563

从数据中可以勾勒出一个典型驾驶者的画像:他大概率是一位三十七、八岁,开着小客车的男性,拥有着较良好的上年度索赔记录($0.88 < 1$)。每一年,他的驾驶里程超过1.6万公里。他的百公里油耗在10升左右,平均速度约为26公里/时。每一天他有2个多小时在路上奔波,其中有十分之一是夜间出行;会做一两次的急加速,也会做一两次的急减速。多数情况下他没有什么行车风险,但一旦发生不测,我们也很难确定他将面临多大的损失。

离散程度最高的是被解释变量——赔偿总金额,该项分布特点确实很特殊,峰度和偏度都极大,最小值为0,其众数亦为0(说明多数人未出险的事实),而最大值则有近36万之多(存在一个偶然的重大车辆事故,会造成巨额损失)。而无赔优调整系数、年龄及百公里油耗等几个解释变量,则有较小的离散程度,体现出一些趋同的特点。

同时也可以关注到车型和性别这两个哑变量,因为其平均值直接反映变量的主体。可以看到样本中小客车占绝大多数,这和预期(私家轿车为主,营运车辆较少)一致。驾车的男性显著多于女性,这也符合常识。

2. 数据分布

表3~6将以几个不同变量作为分类或分段标准,显示整体数据的分布情况。

出险数据分布

表3

年份	#出险	#未出险	出险率	年份	#出险	#未出险	出险率
2010	627	503	55.49%	2013	1573	1843	46.05%
2011	829	798	50.95%	2014	2138	3182	40.19%
2012	1081	1291	45.57%				

车型数据分布

表4

年份	#小客车	#大客车	小客车占比	年份	#小客车	#大客车	小客车占比
2010	1074	56	95.04%	2013	3229	187	94.53%
2011	1546	81	95.02%	2014	4971	349	93.44%
2012	2251	121	94.90%				

性别数据分布

表5

年份	#男性	#女性	男性占比	年份	#男性	#女性	男性占比
2010	884	246	78.23%	2013	2777	639	81.29%
2011	1327	300	81.56%	2014	4513	807	84.83%
2012	1922	450	81.03%				

年龄数据分布

表 6

年份	#40 岁以下	#40 岁以上	40 岁以下占比	年份	#40 岁以下	#40 岁以上	40 岁以下占比
2010	944	186	83.54%	2013	2745	671	80.36%
2011	1372	255	84.33%	2014	4302	1018	80.86%
2012	1967	405	82.93%				

整体上看出险分布,出险率基本稳定在 40%~50% 之间,且逐年下降。根据经验,这个出险水平可能稍高于预期,不过在万级的样本量上也是可以接受的。

五年间,多数变量的数据分布还是比较稳定的。小客车占比始终在 95% 左右,男性占比始终在 80% 上下(尽管该比例是否过高,仍需讨论)。年龄方面,我们依据统计结果及现实规律,将 40 岁作为分段点来考察分布情况。结论也和上文比较一致,即 20~40 岁的年轻驾驶员是主流,占比超过了八成(20 岁以下多不具备驾车资质,故忽略)。

五、结果评价与讨论

(1) 回归结果

将 2014 年的 5320 条数据代入 OLS 回归模型中进行分析。结果见表 7。

OLS 回归结果(2014)

表 7

	Coefficients	标准误差	t Stat	P - value
Intercept	-3421.26	1385.629	-2.46910	0.013577
ClaimAdjustValue	2698.63	818.4327	3.29731	0.000983
MotorType	-778.211	515.0186	-1.51103	0.130839
Gender	-118.665	333.9967	-0.35529	0.722388
Age	-31.1858	15.63927	-1.99407	0.046196
AnnualMileage	-0.03327	0.023142	-1.43781	0.150546
DailyHours	291.478	209.8777	1.38880	0.164952
DailyAcceleration	157.398	40.53142	3.88335	0.000104
DailyDeceleration	-40.3076	67.44728	-0.59762	0.550122
NightDrive	2590.88	1433.183	1.80778	0.070697
AverageSpeed	45.1605	23.25015	1.94237	0.052145
FuelConsumption	241.612	55.57118	4.34779	1.4E-05

在 4 个静态解释变量中,无赔优(NCD)调整系数(ClaimAdjustValue)与赔偿总金额(ClaimAmount)呈正相关关系,其在统计上非常显著;年龄(Age)与索赔总金额呈负相关,其在统计上显著(然而在其他年份年龄均不显著,即显著性并不稳健)。

车型变量(MotorType)与索赔金额呈负相关,性别变量(Gender)亦与索赔金额呈负相关(然而在其他年份,性别与索赔额均呈正相关,也更符合直觉),不过相关关系在统计上均不显著。

在 7 个动态变量中,则有半数以上的显著项。日均急加速次数(DailyAcceleration)和平均百公里油耗(FuelConsumption)两项,与损失金额均具有正相关关系,其在统计上均非常显著;夜间出行比例(NightDrive)和平均速度(AverageSpeed)两项,与损失金额都具有正相关关系,其在统计上都比较显著。

比较惊人的是,年均行驶里程(AnnualMileage)与索赔金额呈负相关,其在统计上不够显著(在有些年份里行驶里程与索赔额呈正相关,然而依然统计不显著)。

日均驾驶时长(DailyHours)与索赔金额呈正相关,日均急减速次数(DailyDeceleration)则与索赔金额呈负相关(在其他年份里急减速同样与索赔额呈负相关,略微反直觉),不过相关关系在统计上均不显著。

(2)分变量的解释

1. 无赔优(NCD)调整系数

传统因子里,无赔优调整系数(以下简称 NCD 系数)是唯一和损失额之间具有显著而稳定的相关关系的。这在直觉上很好理解:NCD 系数越小,反映历史赔款记录越良好,意味着车辆平时行驶表现越出色,其行车风险自然越低,期望损失越小。事实上,由于人的侥幸心理,本存在另一种可能性:上年度无索赔记录所带来的保费折扣,导致人们认为“犯错”成本降低,因而放松了安全驾驶的要求,结果这一年造成了真实损失。不过,相关关系的高度显著有力否定了该猜想,强调了驾驶中的良性循环(或损失的威慑效应)的存在:人们在因 NCD 系数变小而受惠时,倾向于提升(至少是保持)当前的驾驶模式,避免危险驾车行为,此时期望风险不增加。

我们知道损失金额可分解为损失频率和损失强度。而目前的定价模式下,NCD 系数对出险次数或事故率具有显著解释能力。结合平安保险的费率系数表(如表 8),我们可以看到这个线性关系的现实含义,即:每当上年发生赔款的车辆在今年能减少一次赔款,或上年无索赔记录的车辆在今年能够继续保持,那么车辆行驶表现将得到优化,每年可以比上一年减少几百元的期望损失额。这种激励作用是比较明显的。

中国平安费率系数表(2012,NCD 部分)

表 8

费率因子		系数值	说明
以往保险年度 索赔记录 C9	C9a	0.7	连续三年及以上无赔款记录
	C9b	0.8	连续两年无赔款记录
	C9c	0.9	上年无赔款记录
	C9d	1	上年发生二次及以下赔款或首年投保
	C9e	1.1	上年发生三次赔款
	C9f	1.2	上年发生四次赔款
	C9g	1.3	上年发生五次及以上赔款

以上所有结果均凸显了 NCD 系数的重要性。不能忽视的一点是,NCD 系数本就是一个能在宏观上反映人们驾驶行为的综合性因子,这也部分印证了一个假设,即人的驾驶模式具有相对的稳定性,前一阶段驾驶人的行为确实可以用来预测下一阶段驾驶人发生风险驾驶行为的概率大小。

2. 车辆种类、性别与年龄

车辆种类(代码)是一个哑变量。若车辆为小客车,则变量赋值为 0;其余车辆,变量赋值为 1。该车型变量与索赔金额呈负相关的含义是,大客车的索赔风险更小,小客车的索赔风险可能更大。一个很好的解释是,大客车体积大、质量大,相比小客车更厚实,碰撞耐受性更佳,所以期望损失更低。

问题的关键是,该相关关系在统计上均不显著。这意味着,存在一些不能忽视的证据,证明大客车同样可能有更高的风险。为了找出其中的解释,我们决定对车辆分类进行描述性统计,以直接对比二者的异同。表 9 只展示了不同的车辆种类之间差别较大的几方面。

对比发现,小客车确实比大客车有着更高的平均风险。一方面因为体量,另一方面是小客车有着远高于大客车的急加、减速次数。解释是,小客车(包括小轿车)有更多的机动性。人们常开车参加一些非日常性的活动,这时需要走自己不常走的路线前往,会因为不轻车熟路而产生更多不平稳的驾驶行为,如更多的急

加速和急刹车。而大客车多为营运用车,行驶路线往往更为固定,因此行驶模式更为平稳,理应面临更低的行车风险。

我们也注意到,大客车的平均油耗普遍高于小轿车。足够多的证据证明,更高的油耗意味着更高的风险。至此,我们解释了为何大客车也可能具有高风险。这种抵消式的效应,也一定程度上解释了统计不显著的原因。

不同车辆种类的描述性统计(2014,平均值,部分)

表 9

MotorType	ClaimAmount	DailyAcceleration	DailyDeceleration	FuelConsumption
小客车	2023.560698	1.558264291	1.454779003	10.47882788
大客车	1811.825931	0.688223768	1.008932874	13.33035530

自然人性别(代码)是另一个哑变量。若为女性,则变量赋值为 0;若为男性,变量赋值为 1。在多数年份,性别变量与索赔金额呈正相关关系,这才是更符合直觉的。毕竟一般而言,男性可能和鲁莽、草率、粗线条等特征挂钩,而这些特征往往意味着更高的损失风险。

不过问题依然是,性别与损失额的相关关系在统计上不显著。为了寻求解释,我们同样对比男女两性的描述性统计。比较发现,男性在行驶里程和急加、减速次数上显著多于女性,这使得男性面临更大的潜在损失;然而,女性整体上稍显年轻,被认为更加缺乏驾车经验,因此女性也可能具有高风险。这种抵消性的效应,可能部分解释了统计不显著的原因。

最后一个静态变量是年龄。在多数年份,年龄变量始终与损失额呈负相关,但是均不显著。诚然,年轻意味着驾驶经验的缺乏,因而对应更高的风险,不过关系的不显著似乎表明,真实情形更为复杂,特别是对于年龄这样一个常常分段、且在各段都有不同性质的变量。

平安保险的费率系数表(表 10)再次给了我们启示:损失风险与年龄的关系不一定是线性的,也可能是一种 U 型关系——这种关系是符合直觉的。40 岁两侧的描述性统计(未在正文中展示)也部分证明了这一点:年轻人和老年人均可能具有较高的行车风险。

中国平安费率系数表(2012,年龄部分)

表 10

费率因子		系数值	说明
驾驶人年龄 C2	C2a	1.05	年龄 < 25 岁
	C2b	1	25 岁 ≤ 年龄 < 30 岁
	C2c	0.95	30 岁 ≤ 年龄 < 40 岁
	C2d	1	40 岁 ≤ 年龄 < 60 岁
	C2e	1.05	年龄 ≥ 60 岁

当然,考虑到 40 岁以下的驾驶员占据了我们的样本的绝大多数,前半段的相关性质可能反映得更突出,因而在五年中均跑出负相关,也是可以理解的。

3. 年均行驶里程

首代车联网 UBI 产品便是“基于里程的保险”,里程也一直被认为是一个核心动态因子。但五年结果显示,里程与损失额时而正相关、时而负相关,且相关关系均不显著。实际上,在任意里程分段中,里程都没有和赔偿金额呈现出线性关系。这不禁让我们反思变量设置和模型选择的问题。

阅读文献发现,既往研究多数聚焦于驾驶里程与“事故率”的关系,研究其与“赔偿总金额”的关系较少。了解这个情况后,我们将原来的被解释变量换为“是否出险(ClaimOrNot)”这一哑变量,同时换用了 Probit 模型。新的回归结果见表 11。

Probit 回归结果(2014)

表 11

ClaimOrNot	Coef.	Std. Err.	z	P > z
Claimadjustvalue	1.19845	0.1281357	9.35	0.000
Motortype	-0.18767	0.0771765	-2.43	0.015
Gender	-0.09250	0.0490566	-1.89	0.059
Age	-0.00588	0.0023067	-2.55	0.011
Annualmileage	6.70E-06	3.40E-06	1.97	0.049
Dailyhours	-0.00393	0.0059453	-0.13	0.899
Dailyacceleration	0.00997	0.0309397	1.68	0.094
Dailydeceleration	0.00718	0.0099603	0.72	0.471
Nightdrive	0.50871	0.2098265	2.42	0.015
Averagespeed	0.00844	0.0034652	-2.43	0.015
Fuelconsumption	0.00018	0.0082470	0.02	0.983
_cons	-0.97407	0.2079973	-4.68	0.000

我们看到,年均行驶里程和损失频率确实呈显著的正相关关系。在此基础上,我们对里程做分段处理,进一步考察其相关关系,得到图 1。

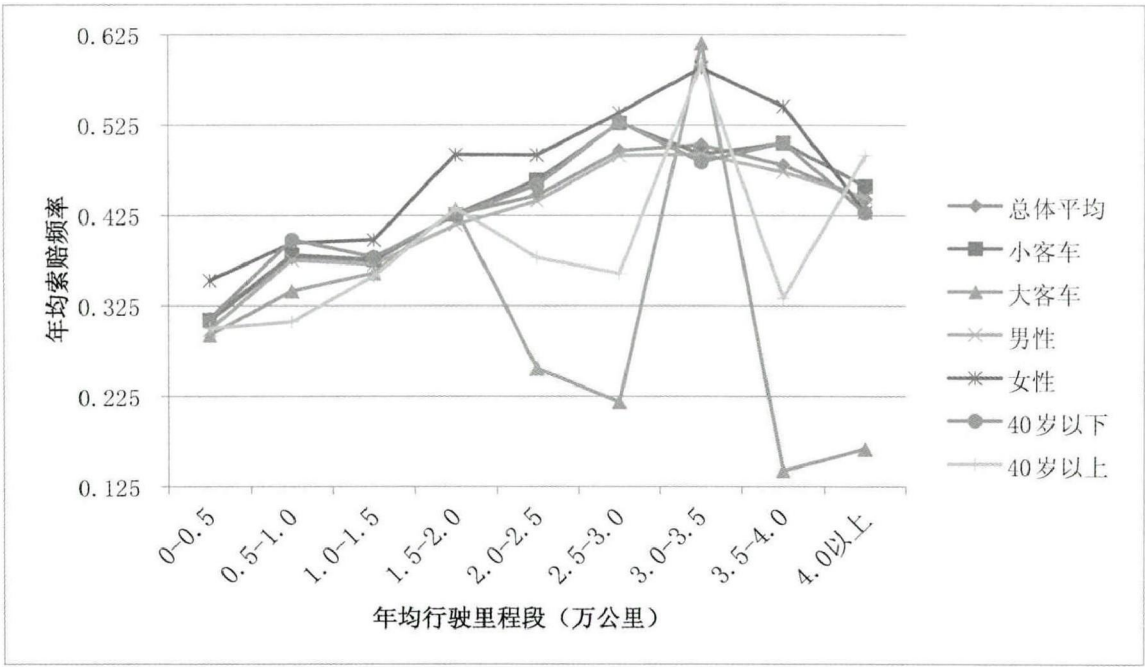


图 1 不同里程段的索赔频率

可以看到,无论从总体上看,还是给定车型、性别或年龄段进行考察,我们都能看到高度类似的递增趋势:随着里程从 0 开始慢慢累积,索赔频率也从 30% 左右逐渐上升,当里程达到 3 万公里时,其对应的索赔率也达到了最高点——50% (由于 3 万公里以上数据量较少,正相关关系没有延续;类似地,由于大客车、女

性和40岁以上人群样本量较小,因此图像波动较大)。至此,我们几乎可以确信,里程虽然不与赔偿总金额直接相关,但是却对事故率有着相当的解释力(里程与索赔频率相关,但与索赔强度相关度低)。对此的解释有很多,例如,里程的增加会让车辆不断折旧、老化,机能在磨损中不断退化;又如,专家认为车辆驾驶中存在着一定的“单位里程事故率”……因此,里程的增加会提高事故率。

另一个问题是,虽然关系显著,但是影响系数却非常小。从图1中也能看到,当里程翻了几番时,损失频率增加还不足一倍。这是什么原因呢?我们猜测,仍然存在一些抵消性的效应。一方面,随着里程增加,驾驶员会积累更多的行车经验,驾驶技术得到提高,同时车子可能磨合得更好,因而可能降低出险率。此外,从结果看,高风险的驾驶员可能会选择少开车,而里程数较多的恰恰是一些行为良好的司机;同时,更安全的车辆,往往开得更远。而且,还有一个不能忽视的现象,即在都市里开车的人常常有更高的事故率,却没有很高的行驶里程。^[11]以上这些情况,均不会使得高里程的司机出现更高的事故率。因此,一定程度上削弱了里程数的解释力,尽管它仍不失为一个不错的定价因子。

4. 日均急加速次数

结果显示日均急加速次数与损失金额具有极其显著的正相关关系。这是一个十分直观的结论:急加速行为使得车辆行驶不平稳,而且加重关键零部件的磨损,因而导致更高的预期风险。

一个有意思的问题是,为什么日均急减速次数呈负相关,且不显著?看上去二者具有很多相似之处,不过我们也找到了一定的解释。

第一,二者性质和用途确实有差别。急减速(急刹车)恰恰经常作为一种规避风险的手段——无论是行车中保持安全距离时,还是在即将碰撞的时刻进行减损时,都有其必要性。相比之下,急加速确实只会增加风险,而不会减少,确实不是一种必要性行为。不妨参考当今流行的车联网安全技术,包括了大量的自动刹车功能,却没有急加速的功能。因此,不难推断,急加速一定趋向于增加风险,而急减速是有可能减少风险的。

第二,根据车联网公司提供的数据,急加速阈值更高(约在 $0.8-1\text{m/s}^2$),急减速阈值低(仅为 $0.001-0.01\text{m/s}^2$)。因此,很多被车联网OBD界定为急减速的行为,可能不构成真正的风险。

5. 夜间出行比例,平均速度及平均百公里油耗

结果告诉我们,夜间出行比例、平均速度、平均百公里油耗均与损失金额具有显著正相关关系。理由也足够直观,此处不再赘述。这几个变量,在未来的保费定价中,理应成为核心因子。

平均百公里油耗值得一提,因为它看似只是一个车属性的变量。然而事实上它同样也是一个驾驶行为的晴雨表。太多经验表明,日常驾驶中一些小的行为,都足以影响油耗,例如超速、低档高速行驶、频繁并线、长时间怠速,乃至换宽胎,都会增加油耗、提高行车风险。

这几个数值变量均体现了良好的线性关系。从含义上讲,每减少1%的夜间出行比例,或平均速度每减少 5m/s ,或者百公里油耗每减少1L,期望损失额都会减少两百元之多,这提供了不错的安全驾驶的激励。

(3) 稳健性检验

第一种办法是分年回归:我们把每一年的数据均代入到模型中进行了回归分析。结果显示,NCD系数、日均急加速次数、夜间出行比例、平均速度和平均百公里油耗在多数年份中均显著,即以上五个因子(一个静态、四个动态)的显著性稳健。

第二种办法是Probit回归:我们将因变量改为哑变量(是否出险),换用Probit模型进行检验。结果参见表5.6。结果显示,NCD系数在2011~2013年依然显著;夜间出行比例在2011年仍然显著,说明这两个因子的显著性稳健。

第三种办法是先分类再回归:我们先控制某一个变量(车型、性别、年龄),在此基础上进行回归分析。我们看到,在2014年,无论在哪一种分类下,NCD系数、日均急加速次数、夜间出行比例、平均速度、平均百公里油耗这五个因子均显著,说明这些因子显著性稳健。

还有一种办法是采用单层决策树分类法:保留原始因变量时,在各年份中,日均急加速次数、平均百公里油耗和夜间驾驶比例均曾被选为第一分类特征,侧面验证了这三个因子显著性的稳健。将因变量改为哑变量时,在所有年份中,NCD 系数均是第一分类特征,侧面验证了其显著性的稳健。

(4) 总体评价

整体而言,我们看到新增的动态变量确实增加了对期望损失的解释力,它们的显著性更好(传统因子则只有一个显著),且具有良好的线性性质,在车联网时代的确是优良的定价因子。

基于以上发现,我们想提出新的保险定价思路。我们希望将车辆种类、性别乃至年龄段这些传统变量作为分类标准,给定这些静态因子,就确定了基础费率(Fixed Charge)。在每个细分下,则由车联网变量(特别是驾驶行为数据)来最终决定保费,即用里程、速度、加速度、夜间比例等动态因子来确定浮动费率(Variable Charge)。通过这种方式,我们能够更好地释放出各类因子的解释力与预测力,进而真正实现有效率的个性化定价。我们期待在不久的将来,能够见证一个双赢的浮动费率激励方案的出生。

(5) 试定价效果

个性化定价的另一个重大意义在于,能够为优质用户节省保费,也帮助保险公司进行产品推广。因此,我们这里采用一定的估计手段进行保费的试定价,并评价其效果是否符合我们的期待(更准确的精算定价,则需在未来的研究工作中进行)。

我们的方法是首先进行两个回归,一个为同时包含新、旧因子的原回归,一个是只包含传统因子的回归。用回归结果中的一系列 β 值分别与各个变量的均值(通过描述性统计得到)分别相乘再相加,可以得到两个损失额的估计值。同时考察承保数据与理赔数据,可以粗略估计出一个“保费损失比”。最后通过损失额估计值与保费损失比的乘积,即可得到两种条件下的保费估计值。

在仅使用传统因子时,保费估计值是 4473.86 元;在使用传统 + 车联网因子时,保费估计值为 4409.62 元,实现了 1.44% 的保费优惠。事实上,五年结果都支持这个结论,每年的优惠率在 0.5% ~ 1.5% 不等,而且整体呈逐年上升趋势(参见表 12)。虽然这个优惠额度还不大,但是已经反映了保险业无限的可能性。只要我们添加动态因子,保费的定价就向更加合理的方向迈进。

保费估计值(2010 ~ 2014 年)

表 12

年份	仅传统因子	传统 + 车联网因子	优惠率
2010	4944.54	4904.01	0.82%
2011	4631.70	4602.11	0.64%
2012	4747.26	4666.80	1.69%
2013	4847.46	4781.99	1.35%
2014	4473.86	4409.62	1.44%

六、结 论

本文主要探究了车联网因子在车辆风险评级中所起到的作用,分析了将驾驶行为变量加入保险公司保费定价的必要性与可行性。论文首先理清了研究思路,选取了合适的变量,形成了研究模型。随后进行了回归分析,并辅以稳健性检验,发现新加入的动态因子(包括急加速、速度、油耗等)对保费的显著解释力,能够弥补传统因子的不足,并且能够实现保费的合理定价。在此基础上,尝试提出了“基础费率(FC) + 浮动费率(VC)”模式,即每个静态因子对应一个定额费用,每个动态因子配备一套费率系数,作为未来的保费定价建议。

本文作为车联网保险领域的初步研究,在很多方面存在可以改进之处。首先,目前所采用的模型还是比

较简单和理想,在后续的研究中,可以采用类似于广义线性模型的高级计量方法,或是机器学习法。其次,本文虽有万级的样本量,但在未来还应继续扩充;同时样本数据出险率偏高,男性比重偏大,可能会产生一定的估计偏差。第三,我们的变量还多少存在着共线性的问题(虽然一般认为,保险定价因子多,共线性难以避免),需要妥善解决;最后,人、车以外的环境因素,未纳入模型中,之后可以考虑补充。

[参考文献]

- [1] 郑苏晋. 车联网与车险风险管理——应用与商业模式研究[M]. 经济科学出版社,2015.
- [2] 王 和. 大数据时代保险变革研究[M]. 中国金融出版社,2014.
- [3] 张 翔. 车联网保险 UBI, 应该怎样管控风险? [EB/OL]. Autolab,2015. <http://www.autolab.cn/2015/12/13/12839/>.
- [4] 韩成丹,赵绰翔,郑苏晋.“行为定价保险”:车险定价与驾驶行为理论[J]. 金融发展研究,2015,(8): 68-74.
- [5] 郁佳敏. 车联网大数据时代汽车保险业的机遇和挑战[J]. 保险研究,2013,(10):89-95.
- [6] 段白鸽,余东发,张连增. 国外车险里程定价理论与实践[J]. 保险研究,2012,(2):72-79.
- [7] Oberholzer M. Strategische implikationen des ubiquitous computing für das nichtleben-geschäft im privat-kundensegment der Assekuranz. Switzerland:Basel University,2003.
- [8] Granhage David, Persson Joahan. A qualitative study about the effects of Transport Information Systems on the insurance premiums for haulage companies. Sweden:Göteborg University,2004.
- [9] Lindberg, Gunnar, et al. "Pay-as-you-speed: Two Field Experiments on Controlling Adverse Selection and Moral Hazard in Traffic Insurance. " Framed Field Experiments 44. 2(2005):262-272.
- [10] Vlad Coroama. The Smart Tachograph-Individual Accounting of Traffic Costs and Its Implications. Pervasive Computing. Springer Berlin Heidelberg,2006:135-152.
- [11] Litman, Todd. "Pay-As-You-Drive Pricing and Insurance Regulatory Objectives. " Journal of Insurance Regulation 23(2005).
- [12] Litman, Todd, and T. Litman. "Distance-based vehicle insurance. " Victoria Transport Policy Institute,2003.
- [13] Litman, Todd. Distance-Based Vehicle Insurance Feasibility, Benefits and Costs: Comprehensive Technical Report, Victoria Transport Policy Institute (www.vtpi.org),2003.
- [14] Ippisch, Tobias. "Telematics Data in Motor Insurance: Creating Value by Understanding the Impact of Accidents on Vehicle Use. " University of St Gallen Business Dissertations (2010).
- [15] Greaves, S., and S. Fifer. "Analysis of a financial incentive to encourage safer driving practices. " Global Positioning System (2014).

Empirical Analysis on Driving Behaviors' Impacts on Auto Risk Rating

SHAN Peng¹, HUA Xiaodong², DENG Yinglu²

(1. Renmin University of China, The School of Finance, Beijing 100872;

2. Tsinghua University, School of Economics and Management, Beijing 100084)

Abstract: This paper mainly studied the role of Internet of Vehicles factor in auto risk rating, and analyzed the necessity and possibility of promoting internationally popular “usage-based insurance” in China. The regression analysis showed that the newly added dynamic factors, including rapid speeding-up, speed, fuel consumption, and so on, were significantly explanative. They could not only make up for the inadequacy of traditional static factors such as sex and age, but also could possibly realize premium reduction. On the basis of this finding, the paper tentatively suggested the “basic rating + floating rating” model for ratemaking.

Key words: Internet of Vehicles; usage-based insurance; ratemaking

[编辑:施 敏]